

# Concept Learning, Classification and Regression

First Assignment for the Lecture Machine Learning



Ulf Krumnack  
Axel Schaffland

Institute of Cognitive Science  
Biologically Motivated Computer Vision

05.04.19

# Outline

---

Organisation

Definitions

Assignments

Resources

# Organisation

---

- Three practice sessions this week
- Assignment 0 will be solved during practice sessions and as homework
- Three task you can choose from
  - Concept learning
  - Classification
  - Regression
- Presentation in tomorrows practice session
  - Each task is presented by one group

# Definitions: Supervised Learning

---

- Learn from example input-output pairs
- Labeled training data: A set of training examples
- Validate and optimize algorithm on validation data
- Test on test data *once* at the end
- Goal: Learn mapping from input data to output label
- Predict output label of new input data
- Example:
  - Object classification
  - Concept learning, classification, regression
- Statistical interpretation: Learn  $p(Y|X)$

# Definitions: Unsupervised Learning

---

- Learning without a teacher
- Self-organisation
- Unlabeled training data
- Goal: Learn relations and grouping in the data
- Examples:
  - Clustering Lego bricks
  - Density estimation, clustering
- Statistical interpretation: Learn something about  $p(X)$

## Definitions: Concept Learning

---

- Acquire the definition of a general category given a sample of positive and negative training examples
- Binary classification
- Indicator function:  $C : x \rightarrow \{0, 1\}$
- Set theory:  $C : x \in A$
- Predicate logic:  $C : x \rightarrow P(x)$
- Example
  - EnjoySport dataset
  - Detection of medical conditions

# Definitions: Classification

---

- Identifying group/class membership
- Predict unordered categorical/discrete variable
- Target function:  $C : x \rightarrow \{c_1, c_2, \dots, c_n\}$
- Examples
  - Handwritten digit classification
  - Object classification

# Definitions: Regression

---

- Estimate a parameter
- Predict value from ordered continuous set (of real numbers)
- Target function:  $R : \mathcal{X} \rightarrow \mathbb{R}$
- Examples
  - House price estimation
  - Diabetes disease progression



# Your Task

---

- Select one of the three presented datasets
- Load a dataset
- Visualize the data
- Train on the data
- Visualize the results
- Focus on visualisation and data handling
- Learning algorithms will be covered later
- Algorithms from **scikit-learn**

# Demo

- Iris Dataset
- Classification of three sub-species: I. setosa, I. versicolor, I. virginica
- Features: Length & Width of Petal & Sepal Leaves

sepal length	sepal width	petal length	petal width	label
5.1	3.5	1.4	0.2	setosa
5.9	3.0	5.1	1.8	virginica
...	...	...	...	...



# Topic I: Concept Learning

---

- Adult Data Set
- Predict whether income exceeds \$50K/yr based on census data.
- Also known as “Census Income” dataset
- Features
  - Age: continuous
  - Workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, ...
  - Education: Bachelors, Some-college, 11th, HS-grad, Prof-school, ...
  - ...

## Topic I: Assignment

---

- Use `csv` to parse the data file
- Pre-process the data with the `LabelEncoder` of sklearn
- Use only continuous attributes (complications with scikit)
- Use only subset of data (else decision tree becomes huge)
- Follow the `Decision Trees Tutorial`
- Use the class `decisionTreeClassifier` and its `fit` function
- Use `graphviz` for visualisation
- Test on the test data set

## Topic II: Classification

---

- Wine Data Set
- Classify the cultivar of grape based on chemical analysis of wine
- Three cultivars/classes, 13 attributes

## Topic II: Assignment

---

- Use `csv` to parse the data file
- Use a similar approach as shown in the demo
- Refer to scikit learn examples [1](#) and [2](#)
- Use all attributes and all classes, opposed to the demo

## Topic III: Regression

---

- Old Faithful Geyser Data

Sample	Eruptions	Waiting
1	3.600	79
2	1.800	54
3	3.333	74
4	2.283	62
5	4.533	85
6	2.883	55
...	...	...



## Topic III: Assignment

---

- Remove the header of the data file
- Use `csv` to parse the data file
- Use `scikits linear regression models`
- `Test` your model with the `live webcam`



# Resources

---

- [scikit-learn tutorials](#)
- [matplotlib pyplot tutorial](#)
- [numpy documentation](#)
- [graphviz](#)
- [pandas](#)
- [xarray](#)